

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-063288

(43)Date of publication of application : 06.03.1998

(51)Int.Cl.

G10L 3/00
G10L 3/00
G10L 3/00
G01C 21/00
G08G 1/0969
G09B 29/10

(21)Application number : 08-241037

(71)Applicant : AQUEOUS RES:KK

(22)Date of filing : 23.08.1996

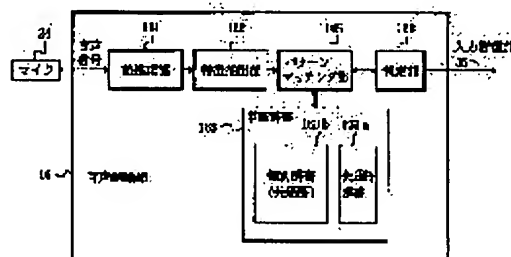
(72)Inventor : HORI KOJI

(54) VOICE RECOGNITION DEVICE

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a voice recognition device in which voices are efficiently recognized by appropriately classifying the contents of a voice dictionary.

SOLUTION: In the device, attention is made to the fact that each word, which it to be recognition object, frequently includes common words (... hotel, for example) representing the meaning contents of each word such as a 'hotel' and a 'bank' in the head of the word, in the middle of the word and the end of the word. Using this fact, the word dictionary is classified into plural individual dictionary 163b for each word, which becomes the recognition object, based on the common words which are owned by the word. Also, common word dictionaries 163 are generated to recognize the common word includes in the inputted voices. Moreover, as pre-recognition, the common words included in the inputted voices are recognized by the pattern matching between the feature section of the latter half of the inputted voices and the dictionaries 163a. After that, inputted voices are recognized by conducting a pattern matching with priority on the individual dictionaries, in which recognized common words are included.



LEGAL STATUS

[Date of request for examination] 17.03.2003

[Date of sending the examiner's decision of rejection] 09.11.2005

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision
of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

* NOTICES *

JPO and INPIT are not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.

2.**** shows the word which can not be translated.

3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] The word dictionary which has the individual dictionary which stored the common language dictionary which stored the standard pattern of a common language, and the standard pattern of two or more words used as the candidate for recognition in the condition which can distinguish the common language contained in the word, A voice input means to input voice, and a feature-extraction means to extract the description about the voice inputted from this voice input means, A common language feature-extraction means to extract the description of a common language part from the description about the input voice extracted with this feature-extraction means, A common language similarity calculation means to compute the similarity of the description extracted with this common language feature-extraction means, and the standard pattern of each common language stored in said common language dictionary, A common language recognition means to recognize the common language contained in the voice inputted from said voice input means from the similarity computed by this common language similarity calculation means, A word dictionary selection means to choose from said word dictionary the standard pattern of a word according to a common language recognized with this common language recognition means, A word similarity calculation means to compute the similarity of the description extracted with said feature-extraction means, and the standard pattern chosen with said word dictionary selection means, The voice recognition unit characterized by providing a judgment means to judge the inputted voice from the similarity computed with this word similarity calculation means.

[Claim 2] The word dictionary which has the individual dictionary which stored the common language dictionary which stored the standard pattern of a common language, and the standard pattern of two or more words used as the candidate for recognition in the condition which can distinguish the common language contained in the word, A voice input means to input voice, and a feature-extraction means to extract the description about the voice inputted from this voice input means, A common language feature-extraction means to extract the description of a common language part from the description about the input voice extracted with this feature-extraction means, A common language similarity calculation means to compute the similarity of the description extracted with this common language feature-extraction means, and the standard pattern of each common language stored in said common language dictionary, A word similarity calculation means to compute the similarity of the description extracted with said feature-extraction means, and the standard pattern chosen with said word dictionary selection means, With a weighting means to perform weighting according to the common language similarity computed with said common language similarity calculation means to the similarity of each word computed with this word similarity calculation means, and this weighting means The voice recognition unit characterized by providing a judgment means to judge the inputted voice from the similarity after carrying out weighting.

[Claim 3] The word dictionary which stored the standard pattern of two or more words used as the candidate for recognition in the condition which can distinguish the genre classified from the contents of the word, The priority table which specified the priority of a genre in order with many storing of the word of the number of alphabetic characters corresponding to each number of

alphabetic characters, A voice input means to input voice, and a number specification means of alphabetic characters to specify the number of alphabetic characters about the voice inputted from this voice input means, A word dictionary selection means to choose from said word dictionary the standard pattern of a word classified into that genre as the priority specified on said priority table according to the number of alphabetic characters specified with this number specification means of alphabetic characters, A feature-extraction means to extract the description about the voice inputted from said voice input means, The voice recognition unit characterized by providing a similarity calculation means to compute the similarity of the description extracted with this feature-extraction means, and the standard pattern chosen with said word dictionary selection means, and a judgment means to judge the inputted voice from the similarity computed with this similarity calculation means.

[Claim 4] The word dictionary which stored the standard pattern of two or more words used as the candidate for recognition in the condition which can distinguish the genre classified from the contents of the word, The priority table which specified the priority of a genre in order with many storing of the word of the number of alphabetic characters corresponding to each number of alphabetic characters, A voice input means to input voice, and a number specification means of alphabetic characters to specify the number of alphabetic characters about the voice inputted from this voice input means, A feature-extraction means to extract the description about the voice inputted from said voice input means, A similarity calculation means to compute the similarity of the description extracted with this feature-extraction means, and the standard pattern chosen with said word dictionary selection means, With a weighting means to perform weighting according to the priority in the number of alphabetic characters specified as the similarity of each word computed with this word similarity calculation means with said number specification means of alphabetic characters specified on said priority table, and this weighting means The voice recognition unit characterized by providing a judgment means to judge the inputted voice from the similarity after carrying out weighting.

[Translation done.]

* NOTICES *

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.**** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention relates to the voice recognition unit which is applied to a voice recognition unit, for example, is used as an input unit in the navigation equipment for cars etc.

[0002]

[Description of the Prior Art] The voice recognition unit which recognizes as language the voice about which human being spoke is put in practical use in the direction of various kinds. It considers using this voice recognition unit as an audio input unit in the case of being put in practical use as an input device directed with voice from the location which left the directions corresponding to the various equipments in works, and carrying out voice input of the destination, the directions information, etc. in the navigation equipment of an automobile. In such a voice recognition unit, since the voice generally inputted is specified, it has the dictionary for speech recognition which extract a spectrum, the time series information on fundamental frequency, etc. as characteristic quantity of input voice, and the pattern is made to correspond to each word, and is stored by analyzing the audio frequency distribution which serves as a candidate for recognition beforehand.

[0003] And if the voice which should be recognized is inputted, comparison collating of the frequency pattern of the inputted voice and the pattern of each word stored in the dictionary for speech recognition will be carried out with pattern matching, and the similarity to each word will be computed. Next, the computed similarity recognizes the highest word (word with the nearest pattern) to be the voice into which it was inputted, and he is trying to output the word. That is, input voice is judged by investigating which word pattern the inputted pattern of frequency distribution of a word bears a strong resemblance to.

[0004] The dictionary for speech recognition used in a voice recognition unit is usually constituted from relation with the matching processing time by 1000 word extent. When the recognition about 1000 or more words is required, it is necessary to prepare two or more dictionaries which divided the word for every group, and to match by changing a dictionary with an application program, and it becomes a problem how the change is performed.

[0005] By the way, the voice recognition unit for mounted information processing indicated by JP,7-64480,A is in the technique which applied the voice recognition unit to the navigation equipment for mount. it be reduction sushi ***** about the probability of the incorrect recognition by the similar word in case comparison collating of the vocabularies which start the contents of a display of the map for navigation equipments register into the voice dictionary in this voice recognition unit, such as the name of a place and a facility name, be carry out and an input word be recognize, even if the vocabulary register into the voice dictionary become extensive, while make the rate of speech recognition of an input word with voice perform quickly efficiently. Therefore, he is trying to determine the group of the voice dictionary used in case an input word is recognized based on the distance over the current position of the car currently called for by navigation equipment, after carrying out the group division of the contents of registration of a voice dictionary according to an area with priority with this navigation

equipment.

[0006]

[Problem(s) to be Solved by the Invention] However, in the voice recognition unit indicated by said official report, since the priority decision index of a voice dictionary is the current position, the count of a change of a voice dictionary increases, so that distance with the position coordinate of the input word of the destination is distant from the current position. Moreover, as long as it was the destination of a vast site which is represented with the name of a place, there could be few counts of a change of a voice dictionary, but when the destination indicated by only the detailed map as shown in a town plan like a store or a home was inputted, there is the need of changing a voice dictionary from the low voice dictionary of whenever [detail / of a map] to the high voice dictionary of whenever [detail] one by one, and retrieval had taken time amount on the contrary.

[0007] The purpose of this invention is by classifying the contents of the voice dictionary appropriately to offer the voice recognition unit which can recognize voice efficiently.

[0008]

[Means for Solving the Problem] The common language dictionary which stored the standard pattern of a common language in invention indicated to claim 1, The word dictionary which has the individual dictionary which stored the standard pattern of two or more words used as the candidate for recognition in the condition which can distinguish the common language contained in the word, A voice input means to input voice, and a feature-extraction means to extract the description about the voice inputted from this voice input means, A common language feature-extraction means to extract the description of a common language part from the description about the input voice extracted with this feature-extraction means, A common language similarity calculation means to compute the similarity of the description extracted with this common language feature-extraction means, and the standard pattern of each common language stored in said common language dictionary, A common language recognition means to recognize the common language contained in the voice inputted from said voice input means from the similarity computed by this common language similarity calculation means, A word dictionary selection means to choose from said word dictionary the standard pattern of a word according to a common language recognized with this common language recognition means, A word similarity calculation means to compute the similarity of the description extracted with said feature-extraction means, and the standard pattern chosen with said word dictionary selection means, From the similarity computed with this word similarity calculation means, a voice recognition unit is made to possess a judgment means to judge the inputted voice, and said purpose is attained. The common language dictionary which stored the standard pattern of a common language in invention according to claim 2, The word dictionary which has the individual dictionary which stored the standard pattern of two or more words used as the candidate for recognition in the condition which can distinguish the common language contained in the word, A voice input means to input voice, and a feature-extraction means to extract the description about the voice inputted from this voice input means, A common language feature-extraction means to extract the description of a common language part from the description about the input voice extracted with this feature-extraction means, A common language similarity calculation means to compute the similarity of the description extracted with this common language feature-extraction means, and the standard pattern of each common language stored in said common language dictionary, A word similarity calculation means to compute the similarity of the description extracted with said feature-extraction means, and the standard pattern chosen with said word dictionary selection means, With a weighting means to perform weighting according to the common language similarity computed with said common language similarity calculation means to the similarity of each word computed with this word similarity calculation means, and this weighting means From the similarity after carrying out weighting, a voice recognition unit is made to possess a judgment means to judge the inputted voice, and said purpose is attained. The word dictionary which stored the standard pattern of two or more words used as the candidate for recognition in the condition which can distinguish the genre classified from the contents of the word in invention according to claim 3, The priority table which

specified the priority of a genre in order with many storing of the word of the number of alphabetic characters corresponding to each number of alphabetic characters, A voice input means to input voice, and a number specification means of alphabetic characters to specify the number of alphabetic characters about the voice inputted from this voice input means, A word dictionary selection means to choose from said word dictionary the standard pattern of a word classified into that genre as the priority specified on said priority table according to the number of alphabetic characters specified with this number specification means of alphabetic characters, A feature-extraction means to extract the description about the voice inputted from said voice input means, A similarity calculation means to compute the similarity of the description extracted with this feature-extraction means, and the standard pattern chosen with said word dictionary selection means, From the similarity computed with this similarity calculation means, a voice recognition unit is made to possess a judgment means to judge the inputted voice, and said purpose is attained. The word dictionary which stored the standard pattern of two or more words used as the candidate for recognition in the condition which can distinguish the genre classified from the contents of the word in invention according to claim 4, The priority table which specified the priority of a genre in order with many storing of the word of the number of alphabetic characters corresponding to each number of alphabetic characters, A voice input means to input voice, and a number specification means of alphabetic characters to specify the number of alphabetic characters about the voice inputted from this voice input means, A feature-extraction means to extract the description about the voice inputted from said voice input means, A similarity calculation means to compute the similarity of the description extracted with this feature-extraction means, and the standard pattern chosen with said word dictionary selection means, With a weighting means to perform weighting according to the priority in the number of alphabetic characters specified as the similarity of each word computed with this word similarity calculation means with said number specification means of alphabetic characters specified on said priority table, and this weighting means From the similarity after carrying out weighting, a voice recognition unit is made to possess a judgment means to judge the inputted voice, and said purpose is attained.

[0009]

[Embodiment of the Invention] Hereafter, the operation gestalt in the voice recognition unit of this invention is explained to a detail with reference to drawing 1 thru/or drawing 3.

(1) In the voice recognition unit of the 1st operation gestalt of ***** of the 1st operation gestalt, each word used as the candidate for recognition notes the point which includes the common word (— flushed — grain ****) showing the semantic content of each word, such as a "hotel" and a "bank", in the initial of the word, a word stem, or the ending in many cases. While carrying out the group division (classification) of the word dictionary to two or more individual dictionaries about each word set as the object of speech recognition based on the common language which that word has using this point, the common language dictionary for recognizing the common language contained in input voice is drawn up. And the common language contained in input voice by pattern matching with a part of input voice and a common language dictionary is recognized as preliminary recognition. Then, pattern matching is preferentially performed from the individual dictionary in which the recognized common language is contained, and input voice is recognized.

[0010] (2) The detail drawing 1 of an operation gestalt expresses the system configuration at the time of applying the voice recognition unit concerning 1 operation gestalt of this invention to navigation equipment. This navigation equipment is equipped with operation part 10. The display 11 containing switch 11b for actuation prepared in the perimeter of display 11a which functions as a touch panel, and this display 11a, and the switch input Management Department 12 which manages the input from the touch panel of this display 11 or switch 11b are connected to this operation part 10.

[0011] There are various switches, such as a switch for performing switch which specifies the menu screen of navigation, switch for adjustment of an air-conditioner, and actuation of an audio, in switch 11b. A push on these switches displays a corresponding menu screen on display 11a.

[0012] operation part 10 -- the current position test section 13, the rate sensor 14, the map information storage section 15, and this operation gestalt -- the speech recognition section 16 to kick and the voice output section 17 are connected. The current position test section 13 is detecting the coordinate data based on the LAT and LONG, and detects the current position which the car has stopped [which has stopped and has current-run]. The GPS (Global Positioning System) receiver 21 which measures the location of a car using a satellite, the beacon receiving set 20 which receives the positional information from the beacon arranged in the road, the bearing sensor 22, and a distance robot 23 are connected to this current position test section 13, and the current position test section 13 measures the current position of a car using the information from these.

[0013] The wheel sensor which computed the amount of displacement of bearing by the bearing sensor 22 arranging gyroscopes, such as a gas rate gyro which detects the angular rate of rotation of the earth magnetism sensor which detects earth magnetism and asks for bearing of a car, and a car, integrates with the angular velocity, and asks for bearing of a car, and an optical fiber gyroscope, and a wheel sensor on either side, and detecting revolution of a car according to the output pulse difference (difference of migration length) is used. Various kinds of approaches, such as that with which a distance robot 23 detects and carries out counting of the rotational frequency of a wheel, or detects acceleration, and it integrates twice, are used. In addition, although location measurement is independently possible for the GPS receiver 21 and the beacon receiving set 20, in the location in which reception by the GPS receiver 21 or the beacon receiving set 20 is impossible, the current position is detected with the bearing sensor 22 and the dead-reckoning navigation which used the both sides of a distance robot 23.

[0014] The map information storage section 15 consists of mass record media, such as CDROM, and a driving gear (driver) of those. Path planning, such as road data required for the path planning to the destination and map data for displaying the path for which it searched on display 11a, and various data required for path guidance are stored in this map information storage section 15. Moreover, the destination data which consist of coordinate data (LAT, LONG) to which the location is indicated to be a name about the various buildings which can be set up as destinations, such as a public facility, a gas station, and a park, or a point are stored in the map information storage section 15. The microphone 24 into which voice is inputted is connected to the speech recognition section 16. The voice output section 17 is equipped with IC26 for voice outputs which outputs voice as an electrical signal, D/A converter 27 which carries out digital to analog conversion of the output of this IC26 for voice outputs, and the amplifier 28 which amplifies the changed analog signal. The loudspeaker 29 is connected to the outgoing end of amplifier 28.

[0015] Operation part 10 realizes each above-mentioned configuration by having CPU (central processing unit), ROM (read only memory), RAM (random access memory), etc., and CPU's making RAM working area and performing the program by which it was stored in ROM or external storage. Namely, the map data reading section 31 by which operation part 10 was connected to the rate sensor 14 and the map information storage section 15, With the map drawing section 32 and the map Management Department 33 which manages the map data reading section 31 and the map drawing section 32 With the screen management section 34 connected to the map drawing section 32 and a display 11, and the input Management Department 35 connected to the switch input Management Department 12 and the speech recognition section 16 It has the voice output Management Department 36 and the communication management section 38 which were connected to IC26 for voice outputs of the voice output section 17 and the map Management Department 33, the screen management section 34, the input Management Department 35, the voice output Management Department 36, and the whole Management Department 37 that manages the communication management section 38. Communication equipment, such as a land mobile radiotelephone which is not illustrated, and PHS, a cellular phone, is connectable at the communication management section 38, and a communication link is managed when performing various communication links, such as multimedia communication, such as facsimile communication besides the usual telephone communication, and personal computer communications, and a communication link by ATIS (Advanced TravelerInformation System).

[0016] Drawing 2 is the block diagram showing the configuration of the speech recognition section 16 in drawing 1. As shown in this drawing, the speech recognition section 16 is equipped with the pretreatment section 161, the feature-extraction section 162, the word dictionary 163, the pattern-matching section 165, and the judgment section 166. The pretreatment section 161 pretreats detection of the voice section, pre-emphasis (high emphasis), noise rejection, etc. to the sound signal after A/D conversion while changing into a digital signal the sound signal inputted from a microphone 24. The feature-extraction section 162 extracts the description about the voice from the sound signal after pretreatment was performed in the pretreatment section 161. Let the description about the extracted voice be the word pattern of the word. Here, the time series information about a spectrum or a cepstrum that the description of a sound signal is acquired by the fast Fourier transform (FFT) is used. This feature-extraction section 162 extracts the description about input voice with various analysis methods, such as a multi-channel band pass filter and linear predictive coding.

[0017] The standard pattern about all words set as the object of speech recognition and this operation gestalt pattern about the common language which has classified each word are stored in the word dictionary 163. This standard pattern is for unspecified speaker recognition, and the description of each word extracted by the same approach as the analytical method of the voice by the feature-extraction section 162 is stored as a standard pattern. It is the destination name which is stored in the contents and the map information storage section 15 of the various assignment keys displayed on the screen of touch panel 11a as a word set as the object of speech recognition, and the various switches of switch 11b and in which a destination setup is possible.

[0018] Drawing 3 expresses an example of the contents of the word dictionary 163 notionally. As shown in this drawing 3, the word dictionary 163 consists of common language dictionary 163a in which the standard pattern of a common language is stored, and two or more individual dictionary 163b in which it is stored for every classification according [the standard pattern of the word for recognition] to a common language. The number of words stored in the dictionary according to each is 1000 or less words. Dictionary 163 according to each b is a bank word dictionary, a hotel word dictionary, a university dictionary, --, other dictionaries (dictionary in which the word which does not contain the common language is stored.), as shown in drawing 3. The recognition word is classified according to the common language contained in each word so that it may say that it does not illustrate. As a word stored in the dictionary according to each, as mentioned above, there is the same word as common languages, such as a name of a prefecture of "kana does not divide" stored in the prefecture dictionary in the case of specifying a name of a prefecture (not shown) and "grain ****" stored in a bank dictionary, etc. as various assignment keys displayed on touch panel 11a. moreover -- as a destination name -- "-- ending -- ** -- grain ****" and a "*****" "check -- it does not obtain like this -- " -- etc. -- it is. The standard pattern of an applicable word and the code information which consists of a sign train corresponding to the word are stored in the dictionary according to each in practice. The code information as the code information corresponding to the contents of an input from the code information on the destination name stored in the map storage section 15, touch panel 11a, etc. that the code information on each word is the same is used.

[0019] The pattern-matching section 165 computes both similarity by comparing the word pattern (description) extracted in the feature-extraction section 162 with the standard pattern stored in the word dictionary 163. Here, pattern matching (comparison) which the pattern-matching section 165 performs has matching by input speech recognition, and matching by the common language recognition for choosing the individual dictionary used with pattern matching of this input speech recognition. In matching for the common language recognition performed first, the word pattern of a common language part is extracted from the word pattern of input voice, the similarity of this and the standard pattern of a common language is computed, and the calculation result is supplied to the judgment section 166. Subsequently, in matching for the input speech recognition performed, an individual dictionary is chosen based on the common language recognition result supplied from the judgment section 166, the similarity of each standard pattern in the individual dictionary and the word pattern extracted in the feature-

extraction section 162 is computed, and the judgment section 166 is supplied.

[0020] The judgment section 166 performs common language recognition which recognizes the common language contained in input voice, and recognition of input voice based on the comparison result of the pattern-matching section 165. In recognition of input voice, the contents of the voice inputted from the microphone 24 are recognized, and the code information corresponding to the contents of recognition (recognition word) is supplied to the input Management Department 35, operation part 10.

[0021] Next, the speech recognition actuation in the voice recognition unit constituted in this way is explained. If the voice which serves as a candidate for recognition from a microphone 24 is inputted into the speech recognition section 16, after changing the inputted audio analog signal into a digital signal and pretreating detection between vocal register, pre-emphasis, noise rejection, etc., in the pretreatment section 161, the sound signal will be supplied to the feature-extraction section 162. In the feature-extraction section 162, the description of the input voice is extracted by analyzing the supplied sound signal. And the extracted description is supplied to the pattern-matching section 165 as a word pattern about the input voice.

[0022] In the pattern-matching section 165, first, the word pattern of a common language part is extracted from the word pattern of input voice, with pattern matching with the standard pattern of each common language of common language dictionary 163a, each similarity is computed and the judgment section 166 is supplied. Here, as a word pattern of the common language part of input voice, the whole word pattern about voice data is equally divided into two, and a part is used the second half. In addition, when the voice section is long, you may make it use the second half parts $1/3$ of a word pattern, or $1/4$. Moreover, the average voice time amount T of all the common languages stored in the common language dictionary is found, and you may make it use the second half part of the word pattern equivalent to this average voice time amount.

[0023] In the judgment section 166, from the similarity to each common language supplied from the pattern-matching section 165, it judges that a common language with the highest similarity is a common language of input voice, and it is supplied to the pattern-matching section 165.

[0024] In the pattern-matching section 165, the individual dictionary of the corresponding common language is chosen out of individual dictionary 163b of the word dictionary 163 based on the judgment result in the judgment section 166. And pattern matching of the word pattern of the input voice already supplied from the feature-extraction section 162 and each standard pattern in the selected individual dictionary is performed, the similarity to each word is computed, and the judgment section 166 is supplied. It sorts to descending of the similarity to each supplied word, and on condition that similarity is over a predetermined threshold, similarity takes out the word of a large high order with the judgment section 166 appointed number N individual picking. And similarity makes the largest word the recognition word to input voice, and similarity supplies the code information corresponding to each word for other words to the input Management Department 35, a control section 10, as the next candidate at descending.

[0025] In addition, from the pattern-matching section 165, in recognition of a common language, similarity supplies a common language high next, and the judgment section 166 requires pattern matching for the second time about other individual dictionaries, when there is no number of words with which the similarity supplied from the pattern-matching section 165 exceeds a predetermined threshold N individual. If the pattern-matching section 165 has this pattern-matching demand for the second time, it will be switched to the individual dictionary of other supplied common languages, will perform pattern matching again, and will supply the similarity about each word to the judgment section 166. The judgment section 166 chooses the high order N individual exceeding a threshold from the similarity already supplied and the similarity newly supplied with pattern matching for the second time, and supplies the code information to the input Management Department 35. When there is no word which exceeds a threshold also by this N individual, the judgment section 166 supplies other common languages to the pattern-matching section 165, and requires pattern matching again until the word exceeding a threshold becomes more than N individual.

[0026] At operation part 10, if the code information corresponding to a recognition word is supplied to the input Management Department, the whole Management Department 37 will check

recognition voice by performing answerback with voice from a loudspeaker 29 through the voice output Management Department 36. Or the predetermined number of the word of N individual corresponding to the supplied code information is displayed on display 11a, and recognition voice is specified because I have a user choose. in addition, the phase which has recognized the common language about answerback with voice -- it is -- for example, -- " -- it feels warm -- alike -- attaching -- ** -- it is -- ? -- " -- you may carry out like. In this case, when a common language is specified by answerback, the above-mentioned pattern-matching processing for the second time to other individual dictionaries is unnecessary.

[0027] As explained above, according to this operation gestalt, each word used as a speech recognition object The common word (-- flushed -- grain ****) showing the semantic content of each word, such as a "hotel" and a "bank", the initial of the word, a word stem, Or while carrying out the group division (classification) of the word dictionary to two or more individual dictionaries about each word set as the object of speech recognition based on the common language which the word has paying attention to the point included in the ending in many cases The common language dictionary for recognizing the common language contained in input voice is drawn up. And since the common language contained in input voice by pattern matching with a part of input voice and a common language dictionary is recognized as preliminary recognition, pattern matching is preferentially performed from the individual dictionary in which the common language recognized after that is contained and it was made to recognize input voice, performing selection and a switch of a voice dictionary appropriately can shorten the recognition time. Moreover, a suitable classification of an individual dictionary and since suitable selection is performed, a recognition rate can be raised.

[0028] Next, the 2nd operation gestalt is explained.

(3) In the voice recognition unit of the 2nd operation gestalt of ***** of the 2nd operation gestalt, each word used as the candidate for recognition can be classified into each genre, such as a hotel, and a bank, a facility, and notes the point that the number of words which consists of the number of the same alphabetic characters in this case differs for every genre. Based on the genre in which that word is contained, the group division (classification) of the word dictionary is carried out to two or more individual dictionaries about each word set as the object of speech recognition using this point. Moreover, the word of each number of alphabetic characters and its number of alphabetic characters creates the priority table which made the priority which put the individual dictionary in order correspond to the order contained. [many] And the number of alphabetic characters of input voice is specified from the voice section, pattern matching is performed in order of the priority of the priority table corresponding to the number of alphabetic characters, and input voice is recognized.

[0029] (4) the 2nd operation gestalt is detailed -- since the system configuration at the time of applying the voice recognition unit of this 2nd operation gestalt to navigation equipment is the same as that of the 1st operation gestalt shown in drawing 1 , omit that explanation. Drawing 4 is the block diagram showing the configuration of the speech recognition section 16 in the 2nd operation gestalt. In addition, the same sign is given to the part which has the function which is common in the speech recognition section 16 of the 1st operation gestalt shown in drawing 2 , the explanation is omitted suitably, and a different functional division functioned and added is explained. In the pretreatment section 161 of this speech recognition section 16, the data (voice duration) of the detected voice section are supplied to the pattern-matching section 165.

[0030] The word dictionary of the speech recognition section 16 is equipped with individual dictionary 163c which classified each word used as the candidate for recognition according to the genre. Although this individual dictionary 163c has many individual dictionary 163b of the 1st operation gestalt classified according to the common language contained in each word, and similar parts, its degree of freedom of a classification at the point that the range of the genre to classify can be specified freely is high, and it differs on following each point. That is, in individual dictionary 163c, since it has classified according to the genre by the semantic content of the word instead of the common language contained in a word, "Hotel New Otani" and "Imperial Hotel" can also be classified into the genre of the same hotel, and distinction by the location of a common language can also be lost. Moreover, "Yasaka Jinja", "Kasuga Taisha", etc. can also

be included in the genre of a shrine, the same common language is not necessarily contained in it, and it may not be in it. Furthermore, it is also possible to summarize the words of the concept been [a concept / it] similar or approximated to the same genre like a shrine and a Buddhist temple. Moreover, since a degree of freedom is in the method of a classification, the genre of a university is also possible also for subdividing a genre like a public university and a private university. Moreover, it is also possible to create genres, such as European food, Japanese-style food, and a teahouse.

[0031] In the pattern-matching section 165 in the 2nd operation gestalt, it has priority table 165b which specified the priority in the case of performing pattern matching to dictionary 163 according to each c as number table of alphabetic characters 165a as which the number of alphabetic characters corresponding to voice duration is specified. About the relation of the audio duration and the number of alphabetic characters of a word in number table of alphabetic characters 165a, both relation is created by measuring beforehand. For example, by two or more persons, about each of two or more words which consist of three characters, I have you utter multiple times every, and distribution of the utterance time amount is measured. Distribution of utterance time amount is similarly measured about the word of other numbers of alphabetic characters. The number table of alphabetic characters is created by extracting a time zone with much [word / each] distribution from this measured value so that it may not overlap.

[0032] Drawing 5 is the explanatory view which expressed the contents of priority table 165b notionally. As shown in this drawing, the sequence of individual dictionary 163c which performs pattern matching is prescribed by the number of alphabetic characters of input voice. By performing pattern matching according to this priority, if the number of alphabetic characters of input voice is 11 characters, the pattern-matching section 165 will perform pattern matching with the department store dictionary of the 1st place of priority first, and will perform pattern matching in order of the bank dictionary of the 2nd less than place, a hotel dictionary, a facility dictionary, and -- if needed after that. Priority table 165b arranges the individual dictionary of each genre in order with many rates (or number) of storing the word of the corresponding number of alphabetic characters. This is created about each genre based on the statistics which investigated the relation between the number of alphabetic characters of an identifier (word), and the rate to the whole (the number of words which has the number of alphabetic characters), as shown in drawing 6.

[0033] Next, actuation of the 2nd operation gestalt constituted in this way is explained. If the voice which serves as a candidate for recognition from a microphone 24 is inputted into the speech recognition section 16, after changing the inputted audio analog signal into a digital signal, while performing predetermined pretreatment and supplying the sound signal after pretreatment to the feature-extraction section 162, in the pretreatment section 161, the voice section data (voice duration) obtained by detection of the voice section will be supplied to the pattern-matching section 165.

[0034] In the feature-extraction section 161, by analyzing the supplied sound signal, the description of the input voice is extracted and the pattern-matching section 165 is supplied as a word pattern about the input voice. In the pattern-matching section 165, while specifying the number of alphabetic characters of the input voice corresponding to the voice duration supplied from the pretreatment section 161 by the time the word pattern was supplied from the feature-extraction section 162 by number table of alphabetic characters 165a, priority table 165b determines the genre of the 1st place of priority to the specified number of alphabetic characters, and individual dictionary 163c about the determined genre is chosen. For example, when it is specified from voice duration that the number of alphabetic characters of input voice is 11 characters, the pattern-matching section 165 chooses the department store dictionary of the 1st place of priority. And if a word pattern is supplied from the feature-extraction section 162, as compared with each standard pattern in the selected individual dictionary, the pattern-matching section 165 will compute the similarity to each word, and will supply it to the judgment section 166.

[0035] It sorts to descending of the similarity to each supplied word, and on condition that similarity is over a predetermined threshold, similarity takes out the word of a large high order

with the judgment section 166 appointed number N individual picking. And similarity makes the largest word the recognition word to input voice, and similarity supplies the code information corresponding to each word for other words to the input Management Department 35, a control section 10, as the next candidate at descending.

[0036] The judgment section 166 requires pattern matching for the second time about other individual dictionaries from the pattern-matching section 165, when there is no number of words with which the similarity supplied from the pattern-matching section 165 exceeds a predetermined threshold N individual. If the pattern-matching section 165 has this demand, priority will switch it to the individual dictionary of a high genre next, and it will perform pattern matching again. For example, when the number of alphabetic characters of the specified input voice is 11 characters, next, priority switches selection to the bank dictionary of the high genre of the 2nd place, and pattern matching is performed. The judgment section 166 chooses the high order N individual exceeding a threshold from the similarity already supplied and the similarity newly supplied with pattern matching for the second time, and supplies the code information to the input Management Department 35. When there is no word which exceeds a threshold also by this N individual, pattern matching is again required until the word exceeding a threshold becomes more than N individual.

[0037] At operation part 10, if the code information corresponding to a recognition word is supplied to the input Management Department, the whole Management Department 37 will check recognition voice by performing answerback with voice from a loudspeaker 29 through the voice output Management Department 36. Or the predetermined number of the word of N individual corresponding to the supplied code information is displayed on display 11a, and recognition voice is specified because I have a user choose.

[0038] In addition, it may be made to perform answerback with voice in the phase which chose the individual dictionary of the genre on a priority table about answerback like the case of the 1st operation gestalt (before performing pattern matching). if it is the case where the number of alphabetic characters of for example, input voice is specified with 11 characters -- first -- the genre of the 1st place of priority -- "department store -- attaching -- ** -- it is -- ? -- " -- it asks like. If a reply is "NO", answerback will be performed about a genre with high priority next one by one. When a common language is specified by answerback, the above-mentioned pattern-matching processing for the second time to other individual dictionaries is unnecessary. Moreover, you may make it display two or more genre names on display 11a about answerback, without being based on voice. The display order in this case is displayed on the order specified on the priority table 165.

[0039] As explained above, according to the voice recognition unit of this 2nd operation gestalt Each word used as the candidate for recognition can be classified into each genre, such as a hotel, and a bank, a facility. In this case, the point that the number of words which consists of the number of the same alphabetic characters differs for every genre is noted. While carrying out the group division (classification) of the word dictionary to two or more individual dictionaries about each word set as the object of speech recognition based on the genre in which the word is contained, the word of each number of alphabetic characters is creating the priority table which arranged the individual dictionary in the order contained. [many] And since the number of alphabetic characters of input voice is specified from the voice section, pattern matching is performed in order of the priority of the priority table corresponding to the number of alphabetic characters and it was made to recognize input voice, performing selection and a switch of a voice dictionary appropriately can shorten the recognition time. Moreover, a suitable classification of an individual dictionary and since suitable selection is performed, a recognition rate can be raised. Moreover, since individual dictionary 163a is classified according to the genre, a degree of freedom is in the method of a classification, and the suitable genre division according to the actual condition, a word for recognition, etc. of a field (speech recognition [in / at this operation gestalt / navigation equipment]) used as the candidate for recognition can be performed.

[0040] It is what explained the operation gestalt of the suitable operation gestalten of this invention with the operation gestalt explained above, and various deformation is possible for this

invention in the range of invention indicated to the claim. For example, with the 1st operation gestalt, although the extract of a common language was extracted from the second half part of the word pattern of input voice, since a common language may exist in the initial of the word like "Hotel New Otani" for example, you may extract from both of a part a part for the first portion, and the second half. For example, when a word pattern is set to one half, it matches with the word pattern for the first portion with each standard pattern of common language dictionary 163a about the both sides of the word pattern of a part in the second half. Moreover, when a word pattern is set to one third, pattern matching is similarly performed about one third the back with one third before a word pattern. In addition, when there is nothing that did not perform the standard pattern and pattern matching of a common language about both the word pattern extracted from a front and after, but performed pattern matching with a next word pattern first, and exceeded the predetermined threshold, it may be made to perform pattern matching with a front word pattern (namely, when matching goes wrong). Furthermore, a common language uses as the word pattern for common language recognition the interstitial segment which may exist in a word stem and divided that word pattern into three equally in this case to the input voice which is the voice section more than predetermined (it is more than the number of predetermined alphabetic characters) like "Green hotel Awaji-cho." When matching by the word pattern of the back and a front goes wrong also in this case, pattern matching about the word pattern of an interstitial segment is performed. In addition, even if it is the same common language, you may make it distinguish an individual dictionary with the location (the initial of the word, a word stem, ending) where a common language exists about the word which contains a common language in the initial of the word or a word stem, although you may make it store in the individual dictionary same as the same classification as the word which contains the common language in the ending. That is, "Imperial Hotel" and "Hotel New Otani" are stored in a different individual dictionary. Thus, input voice can be more certainly recognized by distinguishing with the location where a common language exists.

[0041] Moreover, although the 2nd operation gestalt prescribed the priority of priority table 165b corresponding to the number of alphabetic characters of a word used as the candidate for recognition, you may make it this invention prescribe priority corresponding to the utterance time amount of each word. Namely, about all the words used as the candidate for recognition, the utterance time amount (voice duration) by two or more persons is measured, and an individual dictionary is classified for every mean time of the. For example, voice duration specifies the base of 0.1 seconds, the base of 0.2 seconds, the base of 0.3 seconds, --, the priority of each genre corresponding to the base of m seconds, and --.

[0042] Moreover, for example, although it considered as the configuration which ends pattern matching to other individual dictionaries when the word of the similarity which performs pattern matching one by one in predetermined sequence, and exceeds a predetermined threshold to the dictionary according to each became more than N individual, by this invention, it may be made to perform pattern matching to all individual dictionaries to others with the 1st and 2nd operation gestalten. In this case, weighting according to the total value of the similarity about each combination of an initial-of-the-word vowel and an ending vowel is performed to the similarity obtained as a result of pattern matching to the word of the dictionary according to each. For example, if it is the 1st operation gestalt, weighting according to similarity with the standard pattern of each common language of common language dictionary 163a computed by preliminary recognition will be performed. If it is the 2nd operation gestalt, weighting according to the priority corresponding to the number of alphabetic characters of input voice (the case of the modification of the 2nd operation gestalt utterance time amount) will be performed. In addition, this weighting can be chosen as arbitration about to which range (in the 1st operation gestalt, the similarity in preliminary recognition is what [the] place priority at the 2nd operation gestalt to what position) weighting is carried out. Moreover, it can be chosen as arbitration also about carrying out the multiplication of adding a predetermined value as weighting, or the predetermined multiplier, and an aggregate value and a multiplication value.

[0043] Moreover, although the 1st [which was explained above] and 2nd operation gestalten explained on the assumption that the case where a circuit or a chip etc. which performs pattern

matching is one, you may make it arrange more than one in this invention. For example, you may make it arrange a circuit or a chip for pattern matching of dedication etc. to each of each dictionary of common language dictionary 163a and the individual dictionaries 163b and 163c. In this case, the above-mentioned weighting is carried out. Thus, by considering a circuit or a chip etc. which performs pattern matching as the configuration to arrange [two or more], while recognizing input voice at high speed, a high recognition rate can be acquired.

[0044] Moreover, you may make it store the information on the common language of the word, or the information on a genre in addition to the standard-pattern data and code information on each word, although each word was classified into the individual dictionaries 163b and 163c according to the 1st [which was explained] and 2nd operation gestalten, without carrying out the classification in an individual dictionary. In this case, in the pattern-matching section 165, before performing pattern matching, out of all the words of a word dictionary, the common language supplied from the judgment section 166 or the word corresponding to the genre of priority is selected, and pattern matching is performed after that.

[0045] Moreover, although all the functions of a voice recognition unit were applied to navigation equipment with the operation gestalt explained above, you may make it arrange some or all of a voice recognition unit to other equipments besides navigation equipment in this invention. It is desirable to consider as the information offer station which offers the information about the transit path to the destination etc. by communication link to a car as other equipments. To an information offer station, at least, it is desirable to arrange the speech recognition section 16 whole which includes the pretreatment section 161 and the feature-extraction section 162 although common language dictionary 163a, individual dictionary (common language) 163b or the word dictionary 163 that has individual dictionary (genre) 163c, the pattern-matching section 165, and the judgment section 166 have been arranged to an information offer station, when lessening the equipment configuration by the side of navigation equipment. When the speech recognition section 16 whole has been arranged to the information offer station, voice, such as a destination, is inputted from navigation equipment and this is transmitted to an information offer station from a land mobile radiotelephone etc. through the communication management section 38. In an information offer station, to the voice which received, pretreatment, a feature extraction, pattern matching, and a judgment are performed, and similarity transmits the code information over the recognition word of a high order N individual to navigation equipment by communication link. Which approach explained in the modification about the pattern-matching processing and judgment processing by the information offer station also by the approach explained with said operation gestalt may be used. With navigation equipment, this code information is received through the communication management section 38, and a setup of the destination etc. is performed. In addition, the path planning to the destination is performed and you may make it transmit the information on a search path to navigation equipment in an information offer station based on the destination obtained by speech recognition.

[0046]

[Effect of the Invention] According to this invention, since the contents of the voice dictionary were classified appropriately, voice can be recognized efficiently.

[Translation done.]

* NOTICES *

JPO and INPIT are not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.

2.**** shows the word which can not be translated.

3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] It is a block diagram at the time of applying a voice recognition unit to navigation equipment concerning this invention.

[Drawing 2] It is the block diagram of the speech recognition section in the 1st operation gestalt.

[Drawing 3] It is the explanatory view which expressed notionally an example of the contents of the word dictionary in the 1st operation gestalt.

[Drawing 4] It is the block diagram of the speech recognition section in the 2nd operation gestalt.

[Drawing 5] It is the explanatory view which expressed notionally the contents of the priority table in the 2nd operation gestalt.

[Drawing 6] It is the explanatory view which expressed the relation between the number of alphabetic characters, and the rate to the whole for every genre.

[Description of Notations]

10 Operation Part

11 Display

11a Display

13 Current Position Test Section

15 Map Information Storage Section

16 Speech Recognition Section

161 Pretreatment Section

162 Feature-Extraction Section

163 Word Dictionary

163a Common language dictionary

163b Individual dictionary (common language)

163c Individual dictionary (genre)

165 Pattern-Matching Section

165a The number table of alphabetic characters

165b Priority table

166 Judgment Section

17 Voice Output Section

24 Microphone

33 Map Management Department

34 Screen Management Section

35 Input Management Department

37 Whole Management Department

38 Communication Management Section

[Translation done.]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-63288

(43) 公開日 平成10年(1998) 3月6日

(51) Int.Cl. ⁸	識別記号	庁内整理番号	F I	技術表示箇所
G 1 0 L 3/00	5 3 1		G 1 0 L 3/00	5 3 1 C 5 3 1 G 5 5 1 J 5 7 1 C
G 0 1 C 21/00	5 5 1 5 7 1		G 0 1 C 21/00	H
審査請求 未請求 請求項の数 4 F D (全 12 頁) 最終頁に続く				

(21) 出願番号 特願平8-241037

(22) 出願日 平成8年(1996) 8月23日

(71) 出願人 591261509

株式会社エクス・リサーチ

東京都千代田区外神田 2 丁目19番12号

(72) 発明者 堀 孝二

東京都千代田区外神田 2 丁目19番12号 株式会社エクス・リサーチ内

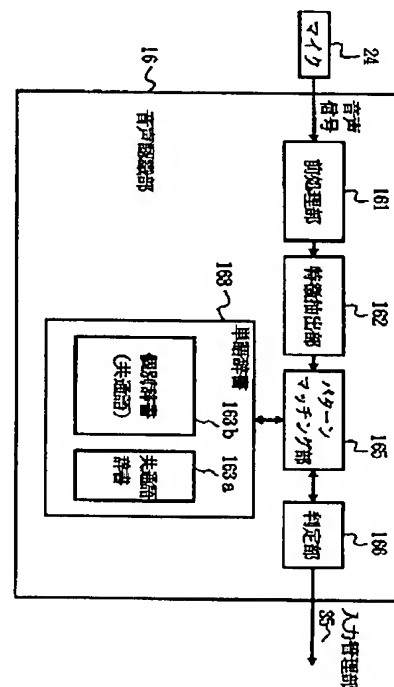
(74) 代理人 弁理士 川井 隆 (外1名)

(54) 【発明の名称】 音声認識装置

(57) 【要約】

【課題】 音声辞書の内容を適切に分類することにより、効率的に音声を認識することが可能な音声認識装置を提供する。

【解決手段】 この音声認識装置では、認識対象となる各単語は、「ホテル」や「銀行」等の各単語の意味内容を表す共通の語（…はてる等）を語頭、語幹、語尾に含んでいることが多い点に着目したものである。この点を利用し、認識対象となる各単語について、その単語が有している共通語に基づいて、単語辞書を複数の個別辞書 163b に分類すると共に、入力音声に含まれる共通語を認識するための共通語辞書 163a を作成しておく。そして、予備認識として、入力音声後半分の特徴部と共通語辞書 163a とのパターンマッチングにより入力音声に含まれている共通語を認識する。その後、認識した共通語が含まれる個別辞書 163c から優先的にパターンマッチングを行い入力音声の認識を行う。



【特許請求の範囲】

【請求項 1】 共通語の標準パターンを格納した共通語辞書と、認識対象となる複数の単語の標準パターンを、その単語に含まれる共通語の区別が可能な状態に格納した個別辞書とを有する単語辞書と、

音声を入力する音声入力手段と、

この音声入力手段から入力された音声についての特徴を抽出する特徴抽出手段と、

この特徴抽出手段で抽出された入力音声についての特徴から、共通語部分の特徴を抽出する共通語特徴抽出手段と、

この共通語特徴抽出手段で抽出された特徴と、前記共通語辞書に格納された各共通語の標準パターンとの類似度を算出する共通語類似度算出手段と、

この共通語類似度算出手段により算出された類似度から、前記音声入力手段から入力された音声に含まれる共通語を認識する共通語認識手段と、

この共通語認識手段で認識された、共通語に応じた単語の標準パターンを前記単語辞書から選択する単語辞書選択手段と、

前記特徴抽出手段で抽出された特徴と、前記単語辞書選択手段で選択された標準パターンとの類似度を算出する単語類似度算出手段と、

この単語類似度算出手段で算出された類似度から、入力された音声を判定する判定手段と、を具備することを特徴とする音声認識装置。

【請求項 2】 共通語の標準パターンを格納した共通語辞書と、認識対象となる複数の単語の標準パターンを、その単語に含まれる共通語の区別が可能な状態に格納した個別辞書とを有する単語辞書と、

音声を入力する音声入力手段と、

この音声入力手段から入力された音声についての特徴を抽出する特徴抽出手段と、

この特徴抽出手段で抽出された入力音声についての特徴から、共通語部分の特徴を抽出する共通語特徴抽出手段と、

この共通語特徴抽出手段で抽出された特徴と、前記共通語辞書に格納された各共通語の標準パターンとの類似度を算出する共通語類似度算出手段と、

前記特徴抽出手段で抽出された特徴と、前記単語辞書選択手段で選択された標準パターンとの類似度を算出する単語類似度算出手段と、

この単語類似度算出手段で算出された各単語の類似度に、前記共通語類似度算出手段で算出された共通語類似度に応じた重み付けを行う重み付け手段と、

この重み付け手段で、重み付けした後の類似度から、入力された音声を判定する判定手段と、を具備することを特徴とする音声認識装置。

【請求項 3】 認識対象となる複数の単語の標準パターンを、その単語の内容から分類されるジャンルの区別が

可能な状態に格納した単語辞書と、

各文字数に対応して、その文字数の単語の格納数が多い順にジャンルの優先順位を規定した優先順位テーブルと、

音声を入力する音声入力手段と、

この音声入力手段から入力された音声についての文字数を特定する文字数特定手段と、

この文字数特定手段で特定された文字数に応じて、前記優先順位テーブルに規定された優先順に、そのジャンルに分類された単語の標準パターンを前記単語辞書から選択する単語辞書選択手段と、

前記音声入力手段から入力された音声についての特徴を抽出する特徴抽出手段と、

この特徴抽出手段で抽出された特徴と、前記単語辞書選択手段で選択された標準パターンとの類似度を算出する類似度算出手段と、

この類似度算出手段で算出された類似度から、入力された音声を判定する判定手段と、を具備することを特徴とする音声認識装置。

【請求項 4】 認識対象となる複数の単語の標準パターンを、その単語の内容から分類されるジャンルの区別が可能な状態に格納した単語辞書と、

各文字数に対応して、その文字数の単語の格納数が多い順にジャンルの優先順位を規定した優先順位テーブルと、

音声を入力する音声入力手段と、

この音声入力手段から入力された音声についての文字数を特定する文字数特定手段と、

前記音声入力手段から入力された音声についての特徴を抽出する特徴抽出手段と、

この特徴抽出手段で抽出された特徴と、前記単語辞書選択手段で選択された標準パターンとの類似度を算出する類似度算出手段と、

この単語類似度算出手段で算出された各単語の類似度に、前記優先順位テーブルに規定された、前記文字数特定手段で特定された文字数における優先順に応じた重み付けを行う重み付け手段と、

この重み付け手段で、重み付けした後の類似度から、入力された音声を判定する判定手段と、を具備することを特徴とする音声認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は音声認識装置に係り、例えば、車両用のナビゲーション装置における入力装置等として使用される音声認識装置に関する。

【0002】

【従来の技術】人間の話した音声を言葉として認識する音声認識装置が各種方面で実用化されている。この音声認識装置は、例えば、工場における各種装置に対応する指示をはなれた場所から音声で指示する入力装置として

10

20

30

40

50

実用化されており、また、自動車のナビゲーション装置において、目的地や指示情報等を音声入力する場合の音声入力装置として用いることが考えられている。このような音声認識装置では、一般に入力された音声を持定するために、予め認識対象となる音声の周波数分布を分析することで、例えば、スペクトルや基本周波数の時系列情報等を入力音声の特徴量として抽出し、そのパターンを各単語に対応させて格納する音声認識用辞書を備えている。

【0003】そして、認識すべき音声が入力されると、入力された音声の周波数パターンと音声認識用辞書に格納された各単語のパターンをパターンマッチングにより比較照合し、各単語に対する類似度を算出する。次に算出された類似度が最も高い単語（パターンが最も近い単語）を、入力された音声であると認識し、その単語を出力するようにしている。つまり、入力された単語の周波数分布のパターンがどの単語パターンに最もよく似ているかを調べることによって、入力音声を判定している。

【0004】音声認識装置において使用される音声認識用辞書は、通常マッチング処理時間との関係から、通常1000単語程度で構成されている。1000以上の単語についての認識が必要な場合には、グループ毎に単語を分けた複数の辞書を用意し、アプリケーションプログラムによって辞書を切り替えて、マッチングを行う必要がある、その切り替えをどのように行うかが問題になる。

【0005】ところで、音声認識装置を車載用のナビゲーション装置に適用した技術に特開平7-64480号公報に記載された、車載情報処理用音声認識装置がある。この音声認識装置では、音声辞書に登録されているナビゲーション装置用の地図の表示内容に係る地名や施設名などの語彙とを比較照合して入力語を認識する際、音声辞書に登録されている語彙が大量になっても、音声による入力語の音声認識率を効率よく迅速に行わせるとともに、類似語による誤認識の確率を低減している。そのために、このナビゲーション装置では、音声辞書の登録内容を地域に応じてグループ分けしたうえで、ナビゲーション装置によって求められている車両の現在位置に対する距離にもとづいて、入力語を認識する際に用いる音声辞書のグループを優先順位をもって決定するようにしている。

【0006】

【発明が解決しようとする課題】しかし、前記公報に記載された音声認識装置では、音声辞書の優先順位決定指標が現在位置であるため、現在位置から目的地の入力語の位置座標との距離が離れているほど音声辞書の切替え回数が増える。また、地名で代表されるような、広大な敷地の目的地であれば音声辞書の切替え回数は少なくてもよいが、商店や個人宅のような市街地図のように詳細な

地図にしか記載されていない目的地を入力した場合は、地図の詳細度の低い音声辞書から詳細度の高い音声辞書へ順次音声辞書を切替える必要性があり、かえって検索に時間を要していた。

【0007】本発明の目的は、音声辞書の内容を適切に分類することにより、効率的に音声を認識することが可能な音声認識装置を提供することにある。

【0008】

【課題を解決するための手段】請求項1に記載した発明では、共通語の標準パターンを格納した共通語辞書と、認識対象となる複数の単語の標準パターンを、その単語に含まれる共通語の区別が可能な状態に格納した個別辞書とを有する単語辞書と、音声を入力する音声入力手段と、この音声入力手段から入力された音声についての特徴を抽出する特徴抽出手段と、この特徴抽出手段で抽出された入力音声についての特徴から、共通語部分の特徴を抽出する共通語特徴抽出手段と、この共通語特徴抽出手段で抽出された特徴と、前記共通語辞書に格納された各共通語の標準パターンとの類似度を算出する共通語類似度算出手段と、この共通語類似度算出手段により算出された類似度から、前記音声入力手段から入力された音声に含まれる共通語を認識する共通語認識手段と、この共通語認識手段で認識された、共通語に応じた単語の標準パターンを前記単語辞書から選択する単語辞書選択手段と、前記特徴抽出手段で抽出された特徴と、前記単語辞書選択手段で選択された標準パターンとの類似度を算出する単語類似度算出手段と、この単語類似度算出手段で算出された類似度から、入力された音声を判定する判定手段と、を音声認識装置に具備させて前記目的を達成する。請求項2に記載の発明では、共通語の標準パターンを格納した共通語辞書と、認識対象となる複数の単語の標準パターンを、その単語に含まれる共通語の区別が可能な状態に格納した個別辞書とを有する単語辞書と、音声を入力する音声入力手段と、この音声入力手段から入力された音声についての特徴を抽出する特徴抽出手段と、この特徴抽出手段で抽出された入力音声についての特徴から、共通語部分の特徴を抽出する共通語特徴抽出手段と、この共通語特徴抽出手段で抽出された特徴と、前記共通語辞書に格納された各共通語の標準パターンとの類似度を算出する共通語類似度算出手段と、前記特徴抽出手段で抽出された特徴と、前記単語辞書選択手段で選択された標準パターンとの類似度を算出する単語類似度算出手段と、この単語類似度算出手段で算出された各単語の類似度に、前記共通語類似度算出手段で算出された共通語類似度に応じた重み付けを行う重み付け手段と、この重み付け手段で、重み付けした後の類似度から、入力された音声を判定する判定手段と、を音声認識装置に具備させて前記目的を達成する。請求項3に記載の発明では、認識対象となる複数の単語の標準パターンを、その単語の内容から分類されるジャンルの区別が可

能な状態に格納した単語辞書と、各文字数に対応して、その文字数の単語の格納数が多い順にジャンルの優先順位を規定した優先順位テーブルと、音声を入力する音声入力手段と、この音声入力手段から入力された音声についての文字数を特定する文字数特定手段と、この文字数特定手段で特定された文字数に応じて、前記優先順位テーブルに規定された優先順に、そのジャンルに分類された単語の標準パターンを前記単語辞書から選択する単語辞書選択手段と、前記音声入力手段から入力された音声についての特徴を抽出する特徴抽出手段と、この特徴抽出手段で抽出された特徴と、前記単語辞書選択手段で選択された標準パターンとの類似度を算出する類似度算出手段と、この類似度算出手段で算出された類似度から、入力された音声を判定する判定手段と、を音声認識装置に具備させて前記目的を達成する。請求項 4 に記載の発明では、認識対象となる複数の単語の標準パターンを、その単語の内容から分類されるジャンルの区別が可能な状態に格納した単語辞書と、各文字数に対応して、その文字数の単語の格納数が多い順にジャンルの優先順位を規定した優先順位テーブルと、音声を入力する音声入力手段と、この音声入力手段から入力された音声についての文字数を特定する文字数特定手段と、前記音声入力手段から入力された音声についての特徴を抽出する特徴抽出手段と、この特徴抽出手段で抽出された特徴と、前記単語辞書選択手段で選択された標準パターンとの類似度を算出する類似度算出手段と、この単語類似度算出手段で算出された各単語の類似度に、前記優先順位テーブルに規定された、前記文字数特定手段で特定された文字数における優先順に応じた重み付けを行う重み付け手段と、この重み付け手段で、重み付けした後の類似度から、入力された音声を判定する判定手段と、を音声認識装置に具備させて前記目的を達成する。

【0009】

【発明の実施の形態】以下、本発明の音声認識装置における実施形態を図 1 ないし図 3 を参照して詳細に説明する。

(1) 第 1 の実施形態の概要

この第 1 の実施形態の音声認識装置では、認識対象となる各単語は、「ホテル」や「銀行」等の各単語の意味内容を表す共通の語（…はてる、…ぎんこう等）を語頭、語幹、または、語尾に含んでいることが多い点に着目したものである。この点を利用して、音声認識の対象となる各単語について、その単語が有している共通語に基づいて、単語辞書を複数の個別辞書にグループ分け（分類）すると共に、入力音声に含まれる共通語を認識するための共通語辞書を作成しておく。そして、予備認識として、入力音声の一部と共通語辞書とのパターンマッチングにより入力音声に含まれている共通語を認識する。その後、認識した共通語が含まれる個別辞書から優先的にパターンマッチングを行い入力音声の認識を行う。

【0010】(2) 実施形態の詳細

図 1 は本発明の一実施形態に係る音声認識装置をナビゲーション装置に適用した場合のシステム構成を表したものである。このナビゲーション装置は、演算部 10 を備えている。この演算部 10 には、タッチパネルとして機能するディスプレイ 11a とこのディスプレイ 11a の周囲に設けられた操作用のスイッチ 11b とを含む表示部 11 と、この表示部 11 のタッチパネルやスイッチ 11b からの入力を管理するスイッチ入力類管理部 12 が接続されている。

【0011】スイッチ 11b には、ナビゲーションのメニュー画面を指定するスイッチ、エアコンの調整用のスイッチ、オーディオの操作を行うためのスイッチ等の各種スイッチがある。これらのスイッチを押すと、対応するメニュー画面がディスプレイ 11a に表示されるようになっている。

【0012】演算部 10 には、現在位置測定部 13 と、速度センサ 14 と、地図情報記憶部 15 と、本実施形態における音声認識部 16 と、音声出力部 17 とが接続されている。現在位置測定部 13 は、緯度と経度による座標データを検出することで、車両が現在走行または停止している現在位置を検出する。この現在位置測定部 13 には、人工衛星を利用して車両の位置を測定する GPS (Global Positioning System) レシーバ 21 と、路上に配置されたビーコンからの位置情報を受信するビーコン受信装置 20 と、方位センサ 22 と、距離センサ 23 とが接続され、現在位置測定部 13 はこれらからの情報を用いて車両の現在位置を測定するようになっている。

【0013】方位センサ 22 は、例えば、地磁気を検出して車両の方位を求める地磁気センサ、車両の回転角速度を検出しその角速度を積分して車両の方位を求めるガスレートジャイロや光ファイバジャイロ等のジャイロ、左右の車輪センサを配置しその出力パルス差（移動距離の差）により車両の旋回を検出することで方位の変位量を算出するようにした車輪センサ、等が使用される。距離センサ 23 は、例えば、車輪の回転数を検出して計数し、または加速度を検出して 2 回積分するもの等の各種の方法が使用される。なお、GPS レシーバ 21 とビーコン受信装置 20 は単独で位置測定が可能であるが、GPS レシーバ 21 やビーコン受信装置 20 による受信が不可能な場所では、方位センサ 22 と距離センサ 23 の双方を用いた推測航法によって現在位置を検出するようになっている。

【0014】地図情報記憶部 15 は、例えば CDROM 等の大容量の記録媒体とその駆動装置（ドライバ）で構成されている。この地図情報記憶部 15 には、目的地までの経路探索に必要な道路データや、探索した経路をディスプレイ 11a に表示するための地図データ等の、経路探索および経路案内に必要な各種データが格納されている。また、地図情報記憶部 15 には、公共施設、ガソ

リンスタンド、公園、等の目的地として設定可能な各種建造物や地点についての名称と、その位置を示す座標データ（緯度、経度）からなる、目的地データが格納されている。音声認識部16には、音声が入力されるマイク24が接続されている。音声出力部17は、音声を電気信号として出力する音声出力用IC26と、この音声出力用IC26の出力をディジタル-アナログ変換するD/Aコンバータ27と、変換されたアナログ信号を増幅するアンプ28とを備えている。アンプ28の出力端にはスピーカ29が接続されている。

【0015】演算部10は、CPU（中央処理装置）、ROM（リード・オンリ・メモリ）、RAM（ランダム・アクセス・メモリ）等を備え、CPUがRAMをワーキングエリアとして、ROMまたは外部記憶装置に格納されたプログラムを実行することによって、上記の各構成を実現するようになっている。すなわち、演算部10は、速度センサ14および地図情報記憶部15に接続された地図データ読込部31と、地図描画部32と、地図データ読込部31および地図描画部32を管理する地図管理部33と、地図描画部32および表示部11に接続された画面管理部34と、スイッチ入力類管理部12および音声認識部16に接続された入力管理部35と、音声出力部17の音声出力用IC26に接続された音声出力管理部36、通信管理部38、および、地図管理部33、画面管理部34、入力管理部35、音声出力管理部36、通信管理部38を管理する全体管理部37とを備えている。通信管理部38には、図示しない自動車電話や、PHS、携帯電話等の通信機器が接続可能になっており、通常の電話通信の他、ファクシミリ通信やパソコン通信等のマルチメディア通信、ATIS（Advanced Traveler Information System）による通信等の各種通信を行う場合に、通信を管理するようになっている。

【0016】図2は、図1における音声認識部16の構成を示すブロック図である。この図に示すように、音声認識部16は、前処理部161、特徴抽出部162、単語辞書163、パターンマッチング部165、および、判定部166を備えている。前処理部161は、マイク24から入力される音声信号をディジタル信号に変換するとともに、A/D変換後の音声信号に対して音声区間の検出、プリエンファシス（高域強調）、雑音除去等の前処理を行うようになっている。特徴抽出部162は、前処理部161で前処理が行われた後の音声信号から、その音声についての特徴を抽出するようになっている。抽出した音声についての特徴は、その単語の単語パターンとされる。ここで、音声信号の特徴は、例えば、高速フーリエ変換（FFT）により得られる、スペクトルやケプストラムについての、時系列情報が使用される。この特徴抽出部162は、多チャンネル・バンドパスフィルタや線形予測分析等の各種分析法によって、入力音声に

ついつつの特徴を抽出するようになっている。

【0017】単語辞書163には、音声認識の対象となるすべての単語についての標準パターンと、各単語を分類している共通語についての本実施形態パターンが格納されている。この標準パターンは、不特定話者認識用のもので、特徴抽出部162による音声の分析方法と同一の方法によって抽出した各単語の特徴が標準パターンとして格納されている。音声認識の対象となる単語としては、タッチパネル11aの画面に表示される各種指定キーとスイッチ11bの各種スイッチの内容、及び、地図情報記憶部15に格納されている目的地設定可能な目的地名等である。

【0018】図3は、単語辞書163の内容の一例を概念的に表したものである。この図3に示すように単語辞書163は、共通語の標準パターンが格納されている共通語辞書163aと、認識対象単語の標準パターンが共通語による分類毎に格納されている複数の個別辞書163bから構成されている。各個別辞書に格納される単語数は1000単語以下になっている。各個別辞書163bは、図3に示すように、銀行単語辞書、ホテル単語辞書、大学辞書、…、その他辞書（共通語を含んでいない単語が格納される辞書。図示しない）、というように、各単語に含まれている共通語により認識単語が分類されている。各個別辞書に格納される単語としては、上述したように、タッチパネル11aに表示される各種指定キーとして、県名を指定する場合の県辞書（図示しない）に格納される「かながわけん」等の県名や、銀行辞書に格納される「ぎんこう」等の共通語と同一の単語等がある。また、目的地名として、「すみともしんこう」、「やさかじんじゃ」、「ひびやこうえん」等がある。各個別辞書には、実際は、該当単語の標準パターンと、その単語に対応する符号列からなるコード情報とが格納されている。各単語のコード情報は、地図記憶部15に格納されている目的地名のコード情報や、タッチパネル11a等からの入力内容に対応したコード情報と同一のコード情報が使用される。

【0019】パターンマッチング部165は、特徴抽出部162で抽出された単語パターン（特徴）と、単語辞書163に格納された標準パターンとを比較することで、両者の類似度を算出するようになっている。ここで、パターンマッチング部165が行うパターンマッチング（比較）は、入力音声認識でのマッチングと、この入力音声認識のパターンマッチングで使用する個別辞書を選択するための共通語認識でのマッチングとがある。最初に行われる、共通語認識のためのマッチングでは、入力音声の単語パターンから共通語部分の単語パターンを抽出し、これと共通語の標準パターンとの類似度を算出し、その算出結果を判定部166に供給するようになっている。次いで行われる、入力音声認識のためのマッチングでは、判定部166から供給される共通語認識結

果に基づいて個別辞書を選択し、その個別辞書内の各標準パターンと特徴抽出部162で抽出された単語パターンとの類似度を算出し、判定部166に供給するようになっている。

【0020】判定部166は、パターンマッチング部165の比較結果に基づいて、入力音声に含まれる共通語を認識する共通語認識と、入力音声の認識とを行う。入力音声の認識では、マイク24から入力された音声の内容を認識し、その認識内容（認識単語）に対応するコード情報を、演算部10の入力管理部35に供給するようになっている。

【0021】次に、このように構成された音声認識装置における音声認識動作について説明する。マイク24から認識対象となる音声が入力されると、前処理部161では、入力された音声のアナログ信号をデジタル信号に変換した後、声区間の検出、プリエンファシス、雑音除去等の前処理を行った後、その音声信号を特徴抽出部162に供給する。特徴抽出部162では、供給された音声信号を分析することで、その入力音声の特徴を抽出する。そして、抽出した特徴を、その入力音声についての単語パターンとして、パターンマッチング部165に供給する。

【0022】パターンマッチング部165では、まず、入力音声の単語パターンから共通語部分の単語パターンを抽出し、共通語辞書163aの各共通語の標準パターンとのパターンマッチングにより、それぞれの類似度を算出して判定部166に供給する。ここで、入力音声の共通語部分の単語パターンとしては、音声データについての単語パターン全体を2等分し、その後半部分を使用する。なお、音声区間が長い場合には、単語パターンの後半部分1/3、または、1/4を使用するようにしてもよい。また、共通語辞書に格納された全共通語の平均音声時間Tを求め、この平均音声時間に相当する単語パターンの後半部分を使用するようにしてもよい。

【0023】判定部166では、パターンマッチング部165から供給される各共通語に対する類似度から、最も類似度が高い共通語を、入力音声の共通語であると判定し、パターンマッチング部165に供給する。

【0024】パターンマッチング部165では、判定部166での判定結果に基づいて、単語辞書163の個別辞書163bの中から、該当する共通語の個別辞書を選択する。そして、既に特徴抽出部162から供給されている入力音声の単語パターンと、選択した個別辞書内の各標準パターンとのパターンマッチングを行い、各単語に対する類似度を算出して判定部166に供給する。判定部166では、供給された各単語に対する類似度の大きい順にソートし、類似度が所定のしきい値を越えていることを条件に、類似度が大きい上位の単語を所定個N個取り出す。そして、類似度が最も大きい単語を入力音声に対する認識単語とし、他の単語を類似度が大き

い順に次候補として、各単語に対応するコード情報を、制御部10の入力管理部35に供給する。

【0025】なお、判定部166は、パターンマッチング部165から供給される類似度が所定のしきい値を越える単語の数がN個無い場合には、パターンマッチング部165に対して、共通語の認識において類似度が次に高い共通語を供給し、他の個別辞書についての再度のパターンマッチングを要求する。パターンマッチング部165は、この再度のパターンマッチング要求があると、供給された他の共通語の個別辞書に切り換えて、再度パターンマッチングを行い、各単語についての類似度を判定部166に供給する。判定部166は、既に供給されている類似度と、再度のパターンマッチングで新たに供給された類似度の中から、しきい値を越える上位N個を選択してそのコード情報を入力管理部35に供給する。これによってもしきい値を越える単語がN個無い場合、判定部166は、しきい値を越える単語がN個以上になるまで、他の共通語をパターンマッチング部165に供給して、再度パターンマッチングを要求する。

【0026】演算部10では、認識単語に対応するコード情報が入力管理部に供給されると、全体管理部37が、音声出力管理部36を介してスピーカ29から音声によるアンサーバックを行うことで、認識音声の確認を行う。または、供給されたコード情報に対応するN個の単語の所定数をディスプレイ11aに表示し、ユーザに選択してもらうことで認識音声を特定する。なお、音声によるアンサーバックについては、共通語を認識した段階で、例えば、「ほてる、についてですか?」のように行ってもよい。この場合、アンサーバックによって、共通語が特定された場合、上記した、他の個別辞書に対する再度のパターンマッチング処理は不要である。

【0027】以上説明したように本実施形態によれば、音声認識対象となる各単語は、「ホテル」や「銀行」等の各単語の意味内容を表す共通の語（…ほてる、…ぎんこう等）を語頭、語幹、または、語尾に含んでいることが多い点に着目し、音声認識の対象となる各単語について、その単語が有している共通語に基づいて、単語辞書を複数の個別辞書にグループ分け（分類）すると共に、入力音声に含まれる共通語を認識するための共通語辞書を作成している。そして、予備認識として、入力音声の一部と共通語辞書とのパターンマッチングにより入力音声に含まれている共通語を認識し、その後、認識した共通語が含まれる個別辞書から優先的にパターンマッチングを行い入力音声の認識を行うようにしたので、音声辞書の選択と切り換えを適切に行うことともに、認識時間を短縮することができる。また、個別辞書の適切な分類と、適切な選択が行われるため、認識率を向上させることができる。

【0028】次に第2の実施形態について説明する。

(3) 第2の実施形態の概要

この第2の実施形態の音声認識装置では、認識対象となる各単語はホテルや銀行、施設等の各ジャンルに分類することができ、この場合、同一文字数からなる単語の数が、各ジャンル毎に異なる点に着目したものである。この点を利用して、音声認識の対象となる各単語について、その単語が含まれるジャンルに基づいて、単語辞書を複数の個別辞書にグループ分け（分類）する。また、各文字数と、その文字数の単語が多く含まれている順に個別辞書を並べた優先順位とを対応させた優先順位テーブルを作成しておく。そして、入力音声の文字数を音声区間から特定し、その文字数に対応する優先順位テーブルの優先順位の順にパターンマッチングを行い、入力音声の認識を行う。

【0029】（4）第2の実施形態の詳細

この第2の実施形態の音声認識装置をナビゲーション装置に適用した場合のシステム構成は、図1に示した第1の実施形態と同一なので、その説明を省略する。図4は、第2の実施形態における音声認識部16の構成を示すブロック図である。なお、図2に示した第1の実施形態の音声認識部16と共通する機能を有する部分には同一の符号を付してその説明を適宜省略し、異なる機能、追加された機能部分について説明する。この音声認識部16の前処理部161では、検出された音声区間のデータ（音声継続時間）をパターンマッチング部165に供給するようになっている。

【0030】音声認識部16の単語辞書は、認識対象となる各単語をジャンル別に分類した個別辞書163cを備えている。この個別辞書163cは、各単語に含まれる共通語によって分類されている第1実施形態の個別辞書163bと類似している部分が多いが、分類するジャンルの範囲を自由に規定することができる点で分類の自由度が高く、次の各点で異なっている。すなわち、個別辞書163cでは、単語に含まれる共通語ではなく単語の意味内容によるジャンルによって分類しているため、「ホテルニューオオタニ」と「帝国ホテル」を同一のホテルのジャンルに分類することもでき、共通語の位置による区別をなくすることもできる。また、神社のジャンルには、「八坂神社」や「春日大社」等を含めることもでき、必ずしも同一の共通語が含まれ無い場合もある。更に、神社と仏閣等のように、類似または近似する概念の単語同士を同一のジャンルに纏めることも可能である。また、分類の仕方に自由度があるため、例えば、大学のジャンルでも、国公立大学と、私立大学というように、ジャンルを細分化することも可能である。また洋食、和食、喫茶店等のジャンルを作成することも可能である。

【0031】第2実施形態におけるパターンマッチング部165では、音声継続時間に対応した文字数が規定されている文字数テーブル165aと、各個別辞書163cに対してパターンマッチングを行う場合の優先順位を規定した優先順位テーブル165bを備えている。文字

数テーブル165aにおける音声の継続時間と単語の文字数との関係については、両者の関係を予め測定することで作成しておく。例えば、複数人により、3文字からなる複数の単語のそれぞれについて、複数回づつ発声してもらい、その発声時間の分布を測定する。同様に他の文字数の単語についても発声時間の分布を測定する。この測定値から、各単語について分布が多い時間帯を、重複しないように抽出することで、文字数テーブルを作成する。

【0032】図5は、優先順位テーブル165bの内容を概念的に表した説明図である。この図に示されるように、入力音声の文字数により、パターンマッチングを行う個別辞書163cの順番が規定されている。パターンマッチング部165は、この優先順位に従ってパターンマッチングを行うようになっており、入力音声の文字数が11文字であれば、まず優先順位第1位のデパート辞書とのパターンマッチングを行い、その後必要に応じて、第2位以下の銀行辞書、ホテル辞書、施設辞書、…の順にパターンマッチングを行うようになっている。優先順位テーブル165bは、該当する文字数の単語を格納している割合（または数）が多い順に各ジャンルの個別辞書を並べたものである。これは、図6に示すように、各ジャンルについて、名前（単語）の文字数と全体（その文字数を有する単語数）に対する割合との関係を調べた統計に基づいて作成される。

【0033】次にこのように構成された第2の実施形態の動作について説明する。マイク24から認識対象となる音声は音声認識部16に入力されると、前処理部161では、入力された音声のアナログ信号をデジタル信号に変換した後、所定の前処理を行い、前処理後の音声信号を特徴抽出部162に供給すると共に、音声区間の検出で得られた音声区間データ（音声継続時間）をパターンマッチング部165に供給する。

【0034】特徴抽出部161では、供給された音声信号を分析することで、その入力音声の特徴を抽出し、その入力音声についての単語パターンとしてパターンマッチング部165に供給する。パターンマッチング部165では、特徴抽出部162から単語パターンが供給されるまでの間に、前処理部161から供給された音声継続時間に対応する入力音声の文字数を文字数テーブル165aにより特定すると共に、特定した文字数に対する優先順位第1位のジャンルを優先順位テーブル165bにより決定し、決定したジャンルについての個別辞書163cを選択しておく。例えば、音声継続時間から入力音声の文字数が11文字であると特定された場合、パターンマッチング部165は、優先順位第1位のデパート辞書を選択しておく。そして、パターンマッチング部165は、特徴抽出部162から単語パターンが供給されると、選択しておいた個別辞書内の各標準パターンと比較し、各単語に対する類似度を算出して判定部166に供

給する。

【0035】判定部166では、供給された各単語に対する類似度の大きい順にソートし、類似度が所定のしきい値を越えていることを条件に、類似度が大きい上位の単語を所定個N個取り出す。そして、類似度が最も大きい単語を入力音声に対する認識単語とし、他の単語を類似度が大きい順に次候補として、各単語に対応するコード情報を、制御部10の入力管理部35に供給する。

【0036】判定部166は、パターンマッチング部165から供給される類似度が、所定のしきい値を越える単語の数がN個無い場合には、パターンマッチング部165に対して他の個別辞書についての再度のパターンマッチングを要求する。パターンマッチング部165は、この要求があると、次に優先順位が高いジャンルの個別辞書に切り換えて、再度パターンマッチングを行う。例えば、特定された入力音声の文字数が11文字である場合には、次に優先順位が高い第2位のジャンルの銀行辞書に選択を切り換えて、パターンマッチングを行う。判定部166は、既に供給されている類似度と、再度のパターンマッチングで新たに供給された類似度の中から、しきい値を越える上位N個を選択してそのコード情報を入力管理部35に供給する。これによってもしきい値を越える単語がN個無い場合には、しきい値を越える単語がN個以上になるまで再度パターンマッチングを要求する。

【0037】演算部10では、認識単語に対応するコード情報が入力管理部に供給されると、全体管理部37が、音声出力管理部36を介してスピーカ29から音声によるアンサーバックを行うことで、認識音声の確認を行う。または、供給されたコード情報に対応するN個の単語の所定数をディスプレイ11aに表示し、ユーザに選択してもらうことで認識音声を特定する。

【0038】なお、第1の実施形態の場合と同様に、アンサーバックについて、優先順位テーブルによるジャンルの個別辞書を選択した段階で（パターンマッチングを行う前に）、音声によるアンサーバックを行うようにしてもよい。例えば、入力音声の文字数が11文字と特定された場合であれば、まず優先順位第1位のジャンルについて、「デパート、についてですか？」のように問い合わせる。回答が「NO」であれば、順次、次に優先順位が高いジャンルについてアンサーバックを行う。アンサーバックによって、共通語が特定された場合、上記した、他の個別辞書に対する再度のパターンマッチング処理は不要である。また、アンサーバックについては、音声によらずに、複数のジャンル名をディスプレイ11aに表示するようにしてもよい。この場合の表示順序は、優先順位テーブル165に規定されている順に表示する。

【0039】以上説明したように、この第2の実施形態

の音声認識装置によれば、認識対象となる各単語はホテルや銀行、施設等の各ジャンルに分類することができ、この場合、同一文字数からなる単語の数が、各ジャンル毎に異なる点に着目し、音声認識の対象となる各単語について、その単語が含まれるジャンルに基づいて、単語辞書を複数の個別辞書にグループ分け（分類）すると共に、各文字数の単語が多く含まれている順に個別辞書を並べた優先順位テーブルを作成している。そして、入力音声の文字数を音声区間から特定し、その文字数に対応する優先順位テーブルの優先順位の順にパターンマッチングを行い、入力音声の認識を行うようにしたので、音声辞書の選択と切り換えを適切に行うことともに、認識時間を短縮することができる。また、個別辞書の適切な分類と、適切な選択が行われるため、認識率を向上させることができる。また、個別辞書163aの分類をジャンルにより行っているため分類の仕方に自由度があり、認識対象となる分野（本実施形態では、ナビゲーション装置における音声認識）の実状や認識対象単語等に応じた適切なジャンル分けを行うことができる。

【0040】以上説明した実施形態では、本発明の好適な実施形態の内の実施形態について説明したもので、本発明は特許請求の範囲に記載した発明の範囲において種々の変形が可能である。例えば、第1の実施形態では、共通語の抽出を入力音声の単語パターンの後半部分から抽出したが、例えば、「ホテルニューオオタニ」のように共通語が語頭に存在する場合があるため、前半部分と後半部分の両者から抽出してもよい。例えば、単語パターンを1/2にした場合、前半部分の単語パターンと、後半部分の単語パターンの双方について、共通語辞書163aの各標準パターンとマッチングする。また、単語パターンを1/3にした場合も同様に、単語パターンの前1/3と、後1/3についてパターンマッチングを行う。なお、前と後から抽出した両単語パターンについて共通語の標準パターンとパターンマッチングを行わず、まず最初に後の単語パターンとのパターンマッチングを行って所定のしきい値を越えたものが無い場合（すなわち、マッチングに失敗した場合）に、前の単語パターンとのパターンマッチングを行うようにしてもよい。更に、「グリーンホテル淡路町」のように、共通語が語幹に存在する場合もあり、この場合には、所定以上の音声区間である（所定文字数以上である）入力音声に対して、その単語パターンを3等分した中間部分を共通語認識用の単語パターンとする。この場合も、後と前の単語パターンによるマッチングに失敗した場合に、中間部分の単語パターンについてのパターンマッチングを行う。なお、語頭や語幹に共通語を含む単語については、語尾にその共通語を含む単語と同一の分類として同一の個別辞書に格納するようにしてもよいが、同一の共通語であっても、共通語が存在する位置（語頭、語幹、語尾）によって個別辞書を区別するようにしてもよい。すなわ

ち、「帝国ホテル」と、「ホテルニューオオタニ」を異なる個別辞書に格納する。このように、共通語が存在する位置によって区別することで、より確実に入力音声を確認することができる。

【0041】また第2の実施形態では、優先順位テーブル165bの優先順位を認識対象となる単語の文字数に対応して規定したが、本発明では、各単語の発声時間に対応して優先順位を規定するようにしてもよい。すなわち、認識対象となる全単語について、複数人による発声時間（音声継続時間）を測定し、その平均時間毎に個別辞書を分類する。例えば、音声継続時間が0.1秒台、0.2秒台、0.3秒台、…、m秒台、…、に対応する各ジャンルの優先順位を規定する。

【0042】また、例えば、第1および第2の実施形態では、各個別辞書に対して、所定の順番で順次パターンマッチングを行い、所定のしきい値を越える類似度の単語がN個以上になった時点で他の個別辞書に対するパターンマッチングを終了する構成としたが、本発明では他に、すべての個別辞書に対するパターンマッチングを行うようにしても良い。この場合、各個別辞書の単語に対するパターンマッチングの結果得られる類似度に対して、語頭母音と語尾母音の各組み合わせについての類似度の合計値に応じた重みづけを行う。例えば、第1の実施形態であれば、予備認識で算出される、共通語辞書163aの各共通語の標準パターンとの類似度に応じた重みづけを行う。第2の実施形態であれば、入力音声の文字数（第2の実施形態の変形例の場合には、発声時間）に対応する優先順位に応じた重みづけを行う。なお、この重み付けをどの範囲まで（第1の実施形態では、予備認識における類似度が何番目まで、第2の実施形態では、優先順位第何位まで）重み付けをするかについては、任意に選択することができる。また、重み付けとして、所定の値を加算するのか、または、所定係数を乗算するのかについて、および、加算値、乗算値についても任意に選択することができる。

【0043】また、以上説明した第1および第2の実施形態では、パターンマッチングを行う回路またはチップ等が1つである場合を前提に説明したが、本発明では、複数配置するようにしても良い。例えば、共通語辞書163a、個別辞書163b、163cの各辞書それぞれに対して、専用のパターンマッチング用の回路又はチップ等を配置するようにしても良い。この場合には、上記した重み付けをする。このように、パターンマッチングを行う回路又はチップ等を複数配置する構成とすることで、入力音声を高速で認識すると共に、高い認識率を得ることができる。

【0044】また、説明した第1および第2の実施形態では、個別辞書163b、163cに各単語を分類したが、個別辞書による分類をすることなく、各単語の標準パターンデータとコード情報に加えて、その単語の共通

語の情報、または、ジャンルの情報を格納するようにしても良い。この場合、パターンマッチング部165では、パターンマッチングを行う前に、単語辞書の全単語の中から、判定部166から供給される共通語、または、優先順位のジャンルに対応する単語をセレクトし、その後にパターンマッチングを行う。

【0045】また、以上説明した実施形態では、音声認識装置の全機能をナビゲーション装置に適用したが、本発明では、音声認識装置の一部、又は全部をナビゲーション装置外の他の装置に配置するようにしても良い。他の装置としては、車両に対して目的地までの走行経路等に関する情報を通信によって提供する、情報提供局とすることが望ましい。情報提供局には、少なくとも、共通語辞書163aと個別辞書（共通語）163b、または、個別辞書（ジャンル）163cを有する単語辞書163と、パターンマッチング部165と、判定部166を配置しておくが、前処理部161、特徴抽出部162を含めた音声認識部16全体を情報提供局に配置しておくことが、ナビゲーション装置側の装置構成を少なくするうえで好ましい。音声認識部16全体を情報提供局に配置した場合、目的地等の音声をナビゲーション装置から入力し、これを通信管理部38を介して自動車電話等から情報提供局に送信する。情報提供局では、受信した音声に対して、前処理、特徴抽出、パターンマッチング、および、判定を行い、類似度が上位N個の認識単語に対するコード情報を、通信によってナビゲーション装置に送信する。情報提供局によるパターンマッチング処理と判定処理については、前記実施形態で説明した方法でも、その変形例で説明したいずれの方法でも良い。ナビゲーション装置では、通信管理部38を介してこのコード情報を受信し、目的地の設定等を行う。なお、情報提供局では、音声認識により得られた目的地に基づいて、その目的地までの経路探索を行い、探索経路の情報をナビゲーション装置に送信するようにしても良い。

【0046】

【発明の効果】本発明によれば、音声辞書の内容を適切に分類したので、効率的に音声を認識することができる。

【図面の簡単な説明】

【図1】本発明に係る、音声認識装置をナビゲーション装置に適用した場合の構成図である。

【図2】第1の実施形態における音声認識部の構成図である。

【図3】第1の実施形態における単語辞書の内容の一例を概念的に表した説明図である。

【図4】第2の実施形態における音声認識部の構成図である。

【図5】第2の実施形態における優先順位テーブルの内容を概念的に表した説明図である。

【図6】文字数と全体に対する割合との関係を各ジャン

17

18

ル毎に表した説明図である。

【符号の説明】

10 演算部

11 表示部

11a ディスプレイ

13 現在位置測定部

15 地図情報記憶部

16 音声認識部

161 前処理部

162 特徴抽出部

163 単語辞書

163a 共通語辞書

163b 個別辞書（共通語）

* 163c 個別辞書（ジャンル）

165 パターンマッチング部

165a 文字数テーブル

165b 優先順位テーブル

166 判定部

17 音声出力部

24 マイク

33 地図管理部

34 画面管理部

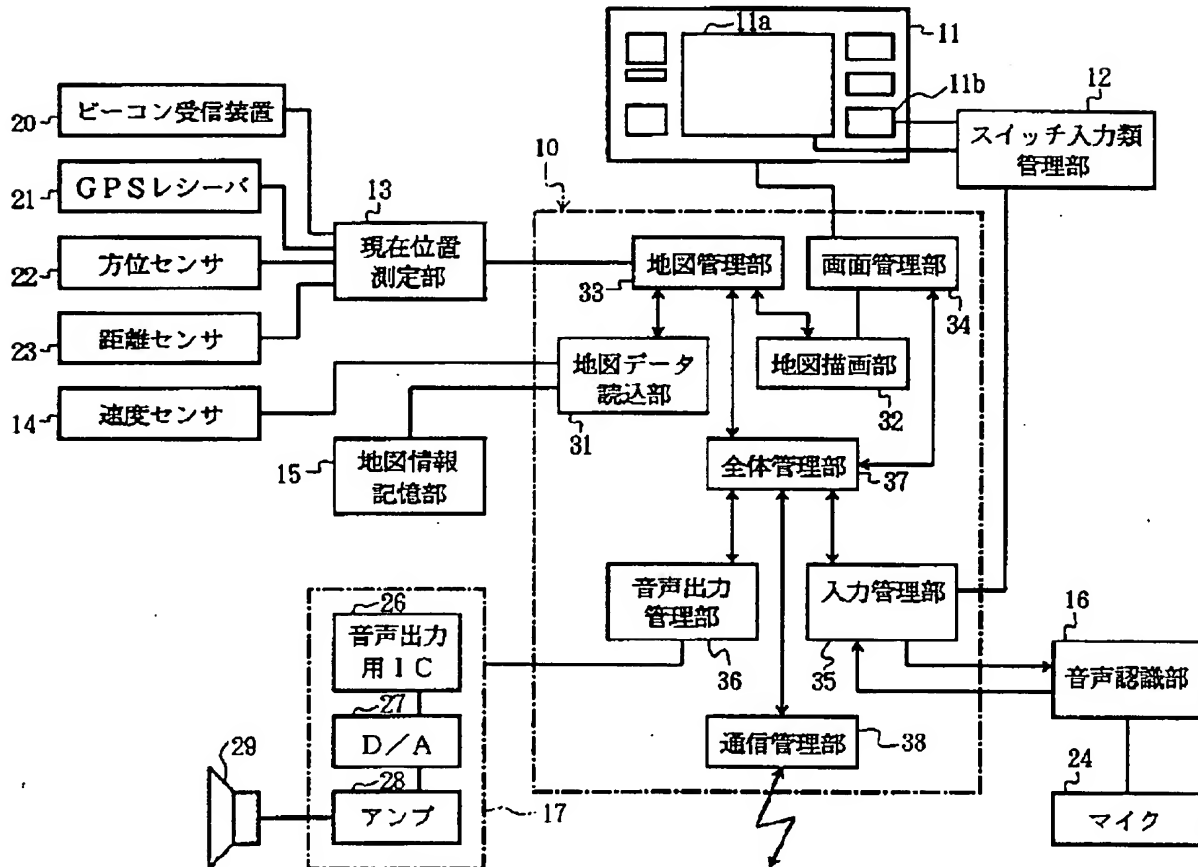
10 35 入力管理部

37 全体管理部

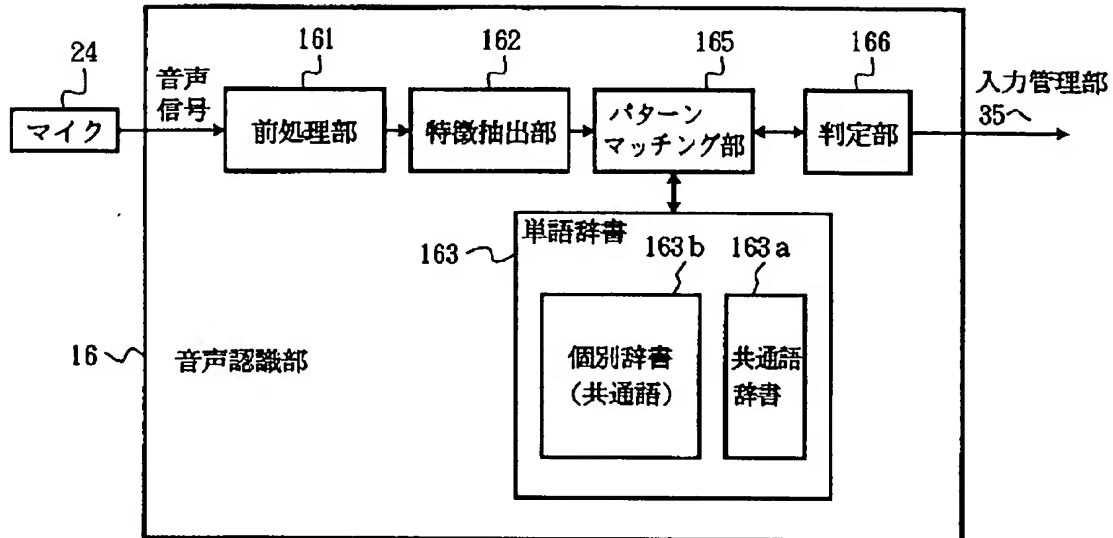
38 通信管理部

*

【図1】



【図2】



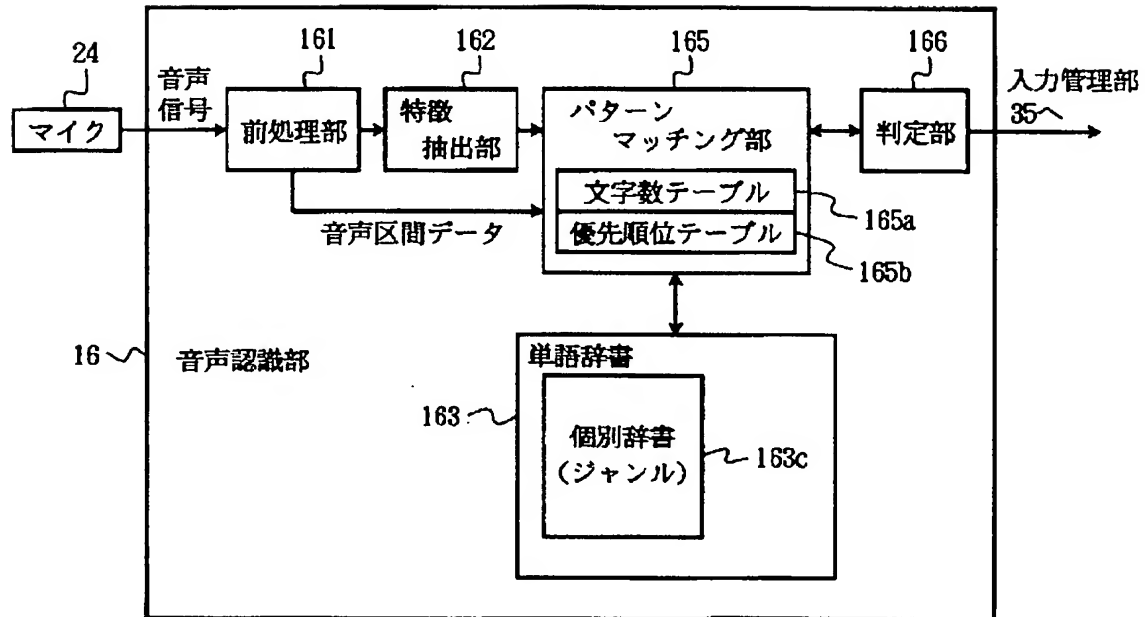
【図3】

163 a					
ぎんこう	ほてる	だいがく	じんじゃ	たいしかん	...
163 b					
銀行辞書	ホテル辞書	大学辞書			
すみともぎんこう	ていこうほてる	わせただいがく			
さくらぎんこう	ばれすほてる	けいおうだいがく			
さんわぎんこう	こくさいかんこうほてる	にほんだいがく			
⋮	⋮	⋮			
神社辞書	大使館辞書	公園辞書			
やさかじんじゃ	あめりかたいしかん	ひびやこうえん			
おおぬまじんじゃ	いざりすたいしかん	きたのまるこうえん			
やすくにじんじゃ	いたりあたishかん	⋮			
⋮	⋮	⋮			
書店辞書	図書館辞書	市役所辞書			
⋮	⋮	⋮			

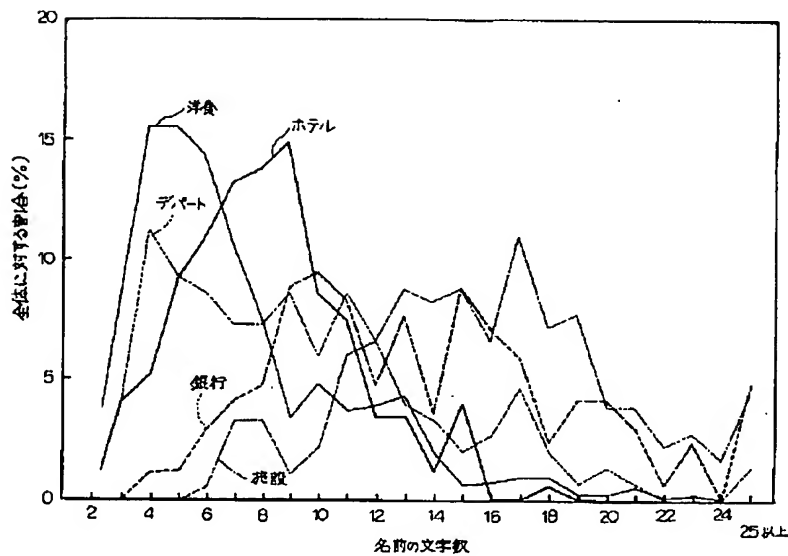
【図5】

文字数	優先順位				
	第1位	第2位	第3位	第4位	...
⋮	⋮	⋮	⋮	⋮	...
11文字	デパート	銀行	ホテル	施設	...
12文字	施設	デパート	銀行	洋食	...
13文字	施設	銀行	洋食	デパート	...
⋮	⋮	⋮	⋮	⋮	...

【図4】



【図6】



フロントページの続き

(51) Int. Cl.⁶

G 0 8 G 1/0969

G 0 9 B 29/10

識別記号

庁内整理番号

F I

G 0 8 G 1/0969

G 0 9 B 29/10

技術表示箇所

A